

Dichte- und Distanzanalyse massiver raumzeitlicher Bewegungsdaten

Stefan BUSCHMANN¹ und Jürgen DÖLLNER²
Hasso-Plattner-Institut, Universität Potsdam

Zusammenfassung. Dieser Beitrag stellt einen Ansatz zur interaktiven Analyse massiver raumzeitlicher Bewegungsdaten vor. Der Ansatz ermöglicht es, auf Bewegungstrajektorien basierende Daten interaktiv zu visualisieren und mit flächenbasierten Verfahren auf Basis von Dichte- und Distanzkarten zu analysieren. Ziel ist es, dadurch unterschiedliche raumzeitliche Phänomene in massiven Bewegungsdaten aufspüren, extrahieren und untersuchen zu können. Diese Verfahren stellen generische Basisfunktionen für die Auswertung massiver raumzeitlicher Bewegungsdaten, z. B. für entscheidungsunterstützende Geoinformationssysteme, bereit. Anwendungsfelder umfassen die Echtzeit-Bewegungsüberwachung und die Auswertung von Bewegungsdaten, z. B. im Straßen-, Schiffs- oder Flugverkehr.

Keywords. Raumzeitliche Daten, Explorative Visualisierung, Visual Analytics

1. Einleitung

In einer wachsenden Zahl von Anwendungsbereichen werden sensorgestützt Daten über die Bewegung von Objekten erfasst, z. B. im Straßen-, Schiffs- oder Flugverkehr. Diese Daten sind, vor allem wenn sie in Echtzeit erfasst werden, im Allgemeinen durch ihre große Menge gekennzeichnet und erfordern daher spezielle Verfahren zur Filterung, Visualisierung, Analyse und Exploration der darin enthaltenen raumzeitlichen Phänomene.

In diesem Beitrag stellen wir einen Ansatz zur analytischen und explorativen Visualisierung massiver Bewegungsdaten vor, der auf generischen Dichtekarten beruht. Die vorgestellten Methoden stellen damit eine Ergänzung zu anderen Ansätzen, z. B. Clusteranalysen [1] [2] in raumzeitlichen Daten oder der Dichteanalyse auf Basis von Faltungskernen [3] [4], dar. Ziel ist es dabei, Phänomene in großen Mengen raumzeitlicher Daten erkennbar und explorierbar [5] zu machen, um so raumzeitliche Phänomene interaktiv untersuchen zu können.

Dieser Ansatz kann beispielsweise als Baustein für die Entwicklung entscheidungsunterstützender Systeme, z. B. bei der Flugsicherung oder beim Verkehrsmonitoring, dienen und wird daher auf seine Eignung innerhalb von Echtzeit-Analyse- und Visualisierungssystemen hin untersucht. Die Implementierung der Techniken erfolgt auf Basis moderner GPU (Graphics Processing Unit) basierter

¹ Stefan.Buschmann@hpi.uni-potsdam.de

² Juergen.Doellner@hpi.uni-potsdam.de

Systeme, die durch ihre massive Parallelisierung eine echtzeitfähige Berechnung der vorgestellten Visualisierungsmethoden ermöglichen.

2. Raumzeitliche Bewegungsdaten

Raumzeitliche Bewegungsdaten unterscheiden sich von allgemeinen raumzeitlichen Daten dadurch, dass sie Bewegungen von Objekten im physikalischen Raum beschreiben. Dabei handelt es sich um kontinuierliche Bewegungen von Objekten, die sich in einer bestimmten Zeit von einer Position A zu einer Position B bewegen und dabei einen ununterbrochenen Weg zurücklegen, der durch Linien- und Kurvensegmente approximiert werden kann. Im Gegensatz hierzu können allgemeine raumzeitliche Daten z. B. punktuelle Ereignisse beschreiben, welche zwar einen räumlichen und zeitlichen Bezug haben, dabei jedoch im eigentlichen Sinne keine Bewegung beschreiben. Die Repräsentation von Bewegungsdaten erfolgt meist durch eine diskrete Abtastung der Bewegung, z. B. durch in festgelegten Abständen gemessene Wegpunkte oder durch kontinuierliche Kurvensegmente (z. B. B-Splines).

Die Bewegungstrajektorien dienen als Datengrundlage für den hier vorgestellten Visualisierungs- und Analyseansatz. Beispielsweise liefern die Positionsaufzeichnungen einer Fahrzeugflotte oder Aufzeichnungen über die Bewegungen von Flugzeugen im Luftraum massive raumzeitliche Bewegungsdaten.

3. Dichteanalyse

Als ersten Visualisierungsansatz betrachten wir die Häufigkeit des Auftretens von Objekten bezogen auf ihre geografische Position. Das Ergebnis lässt sich als Dichtekarte interpretieren, also als zweidimensionales Raster, in dem jeder Rasterzelle die Anzahl an Objekten zugewiesen wird, welche diese im betrachteten Zeitraum passiert haben. Dieser Ansatz dient dazu, die Bewegungsdichte zu analysieren und in Form einer Belastungskarte darzustellen. So können Belastungssituationen in einzelnen geografischen Bereichen untersucht werden.

Grafisch lässt sich eine solche Dichtekarte beispielsweise als Heatmap [6] darstellen, also durch Abbildung der Dichtewerte auf eine Farbe, welche dem Benutzer in Form einer Karte bzw. eines Kartenlayers dargestellt werden kann. Dabei wird eine Farbcodierung verwendet, die jeder Dichteinformation eine Farbe zuweist. Hierzu wird eine Transferfunktion definiert, welche aus dem Wertebereich der Dichtekarte auf den Farbbereich abbildet.

Zur Interpretation einer Dichtekarte ist es notwendig, die Transferfunktion auf Basis der zugrunde liegenden Daten so zu wählen, dass ein aussagekräftiges visuelles Ergebnis entsteht. So müssen die einzelnen Belastungswerte beispielsweise in Relation gesetzt werden zu den vorgefundenen Minimal- und Maximalwerten, um lokal hohe und geringe Belastungen visuell differenzierbar darzustellen. Dies zeigt sich insbesondere im Zusammenspiel mit der temporalen Exploration der Daten: Über einen kurzen Zeitraum betrachtet ist die aggregierte maximale Anzahl von Objekten deutlich niedriger, als wenn ein längerer Zeitraum betrachtet wird. Um dennoch lokale Minima und Maxima darstellen zu können, muss die Farbskala entsprechend angepasst werden. Die Wahl der Farbskala und die dazugehörige Transferfunktion für die

Auswahl der Farbe müssen daher dynamisch an die gewählten Eingabedaten angepasst werden.

4. Raumzeitliche Distanzanalyse

Ein anderes Kriterium zur Bewertung von Belastungssituationen stellt die Distanz zwischen Objekten dar. Eine geringe Distanz zwischen vielen Objekten kann ein Indiz für ein hohes Verkehrsaufkommen sein, da sich die einzelnen Objekte dichter aneinander vorbei bewegen müssen. Um dieses Phänomen zu untersuchen, kann wiederum eine Dichtekarte verwendet werden, in der anstelle der Anzahl die Distanz zwischen den Objekten gespeichert wird. Distanz wird dabei räumlich und zeitlich definiert:

- **Räumliche Distanz:** Die räumliche Distanz zwischen zwei Objekten wird über die geografische Distanz zwischen ihren jeweiligen Positionen ermittelt. In vielen Anwendungsfällen reicht hier als Näherung die euklidische Distanz aus, sofern der betrachtete geografische Bereich eine bestimmte Größe nicht übersteigt. Beispielsweise erfüllt die Betrachtung von Flugbewegungen im An- und Abflugbereich eines Flughafens (ca. 100km x 100km) diese Bedingung.
- **Raumzeitliche Distanz:** Die raumzeitliche Distanz ist eine Funktion, die für zwei Objekte deren räumliche Distanz zu einem Zeitpunkt bestimmt. Ihr Definitionsbereich ist dabei die Schnittmenge der Definitionsbereiche beider Objekte, enthält also alle Zeitpunkte, an denen beide Objekte gleichzeitig definiert sind.
- **Distanz-Dichte:** Für die raumzeitliche Betrachtung der Distanz in Form einer Dichtekarte wird für jedes Rasterfeld die Distanz zwischen allen Paaren von Objekten betrachtet, deren Verbindungslinie das jeweilige Feld kreuzen. Hierdurch werden alle relevanten Distanzen, die zwischen je zwei Objekten innerhalb des betrachteten Zeitraumes auftraten, ermittelt. Aus diesen Distanzwerten wird nun mit einer für die jeweilige Fragestellung geeigneten Operation (wie z. B. Minimal- oder Mittelwert) der Dichtewert bestimmt.

Die erzeugte Distanzkarte (Abb. 1) stellt die Distanz zwischen Objekten räumlich dar. Dabei beschreibt der Wert einer Rasterzelle die Distanz zwischen den Objekten in der Umgebung der Zelle, wodurch ein räumliches Bild der Distanzen entsteht.

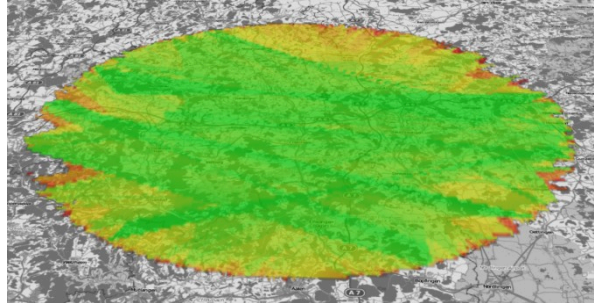


Abbildung 1: Distanzkarte von Flugzeugen im Radarbereich um einen Flughafen.

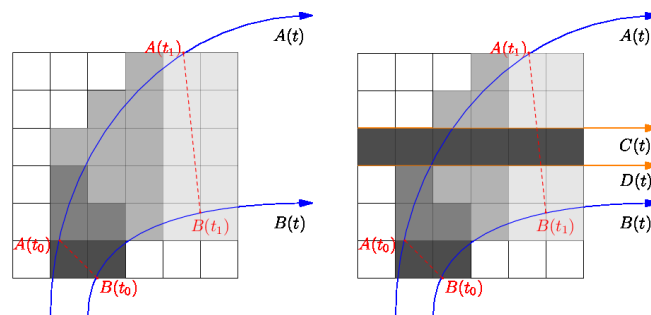


Abbildung 2: Konstruktion einer Distanzkarte unter Verwendung des Minimums als Operation. Das linke Bild zeigt die Distanzkarte für zwei Trajektorien A und B , das rechte Bild enthält zwei weitere Trajektorien C und D , welche keine zeitlichen Überschneidungen mit A und B besitzen.

Zur Konstruktion der Distanzkarte werden alle Bewegungstrajektorien über den gewählten Zeitraum betrachtet und die raumzeitliche Distanz zwischen je zwei Trajektorien bestimmt. Diese Distanzen werden jeweils in der Verbindungsgerade zwischen den aktuellen Positionen der beiden Objekte unter Verwendung der gewählten Operation auf der Karte abgebildet (Abb. 2).

Im Gegensatz zur Objektanzahl, welche relativ zu einem Maximalwert betrachtet werden muss, stellt die Distanz zwischen Objekten eine absolute Größe dar. Aus diesem Grund ist es bei der Berechnung von Distanzkarten nicht notwendig, die Farbskala der Distanzkarte dynamisch mit der Anzahl an Eingabedaten anzupassen. Jedoch kann eine Anpassung der Skala im Zusammenhang mit einer räumlichen Exploration sinnvoll sein, um die Größenverhältnisse der Distanzkarte denen des betrachteten Kartenausschnitts anzupassen. Ob eine solche Anpassung sinnvoll ist, sollte jedoch in Abhängigkeit von der zu untersuchenden Fragestellung entschieden werden.

Ebenfalls abhängig von der jeweiligen Fragestellung ist die Wahl der Operation, welche für die Distanzkarte verwendet wird. Beispielsweise kann die minimale Distanz zwischen Fahrzeugen als Hinweis auf lokale Engpässe in einer Verkehrssituation dienen und dazu verwendet werden, lokale Hindernisse (auch zeitlich beschränkte) zu erkennen. Die mittlere Distanz kann dagegen besser geeignet sein, um als Indiz für die Bewertung der Gesamtsituation zu dienen, da sich in einer angespannten Verkehrssituation die mittlere Distanz aller Objekte verringern dürfte.

5. Interaktive raumzeitliche Exploration

5.1. Räumliche Exploration

Zur Verwendung der vorgestellten Visualisierungsmethoden empfiehlt sich die Einbindung der Verfahren innerhalb eines interaktiven Geovisualisierungssystems, wie z. B. eines virtuellen Globus oder interaktiver Atlanten. So können Dichtekarten beispielsweise als zusätzliche Kartenlayer in ein Geoinformationssystem (GIS) eingebunden werden und fügen sich damit nahtlos in die Interaktions- und Explorationslogik dieses Systems ein.

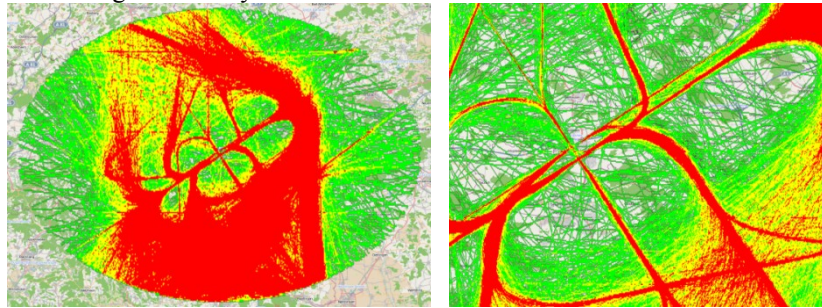


Abbildung 1: Generierte Dichtekarte in zwei unterschiedlichen Zoomstufen.

Da die Berechnung der Dichte- und Distanzkarten in Echtzeit erfolgt, können diese Karten während der räumlichen Exploration durch den Benutzer automatisch neu erstellt und somit an die geänderten Visualisierungsparameter angepasst werden. So lassen sich beispielsweise Größe und Auflösung der Dichte- und Distanzkarten automatisch anpassen, sobald ein anderer Kartenausschnitt gewählt oder die Zoomstufe verändert wurde, wodurch z. B. ein genaueres Bild des gewählten Ausschnitts mit einer höheren Auflösung berechnet wird. Die zugrunde liegenden Eingabedaten für die Berechnung der Dichtekarte bleiben dabei unverändert, durch ein höher aufgelöstes Raster wird der gewählte Bereich jedoch feinstufiger analysiert (Abb. 3).

Hierdurch können die Visualisierungsverfahren auch gut in bestehende Methoden der Kachelung und Zwischenspeicherung eingebunden werden, welche vom verwendeten Geovisualisierungssystem zur Verfügung gestellt und verwaltet werden. Das System fragt dabei neue Kartenausschnitte an, welche dann automatisch mithilfe der vorgestellten Methoden berechnet und zur Verfügung gestellt werden können. Die Verwaltung und Zwischenspeicherung der Kartenausschnitte erfolgt dabei durch das Geovisualisierungssystem selbst.

5.2. Zeitliche Exploration

Die Methoden für die Berechnung von Dichte- und Distanzkarten basieren auf der Aggregation von Bewegungsdaten über einen definierten Zeitraum. Dadurch werden zeitliche Verläufe nicht explizit dargestellt, sondern fließen indirekt durch die Akkumulation von Daten ein. Die Wahl des betrachteten Zeitraumes ist daher entscheidend für das Ergebnis und dessen Interpretation. Durch eine Veränderung des

betrachteten Zeitraums wird dabei die Auswahl der Daten aus der Menge der Eingabedaten verändert, die für die Aggregation und Visualisierung herangezogen werden. Dies hat eine direkte Auswirkung auf das Ergebnis der vorgestellten Visualisierungsverfahren und kann so zur Analyse von zeitlichen Veränderungen verwendet werden (Abb. 4).

Für eine temporale Exploration der Daten ist es daher notwendig, dass der betrachtete Zeitraum interaktiv durch den Anwender verändert werden kann. Hierfür ist auf Seite des Anwenders Expertenwissen über die Anwendungsdomäne erforderlich, um durch eine gezielte Exploration vorliegende Fragestellungen untersuchen und das Ergebnis entsprechend interpretieren zu können.

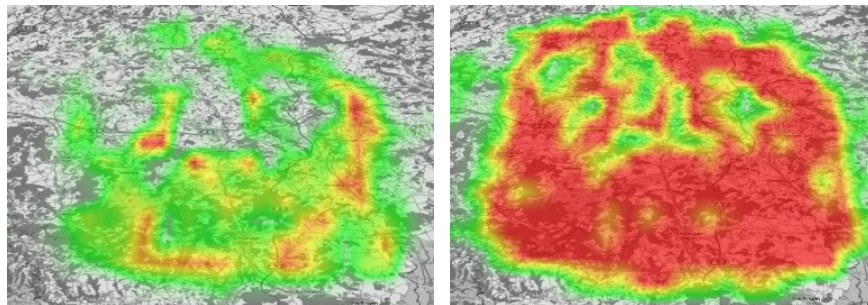


Abbildung 2: Zeitliche Exploration einer Dichtekarte. Das linke Bild zeigt ein Zeitfenster von 00:00 Uhr bis 08:00 Uhr, im rechten Bild wird der Zeitraum von 00:00 Uhr bis 24:00 Uhr betrachtet.

Die interaktive zeitliche Exploration der Daten erfolgt über zwei Parameter, die für jede Visualisierung gewählt und interaktiv verändert werden können: Den globalen Zeitraum und ein Zeitfenster pro Tag. Der Zeitraum definiert den Anfang und das Ende des Zeitintervalls in Tagen. Das Zeitfenster beschreibt den betrachteten Zeitraum pro Tag. So können z. B. nur Tag oder nur Nacht betrachtet und miteinander verglichen werden.

Im Szenario der Verkehrsüberwachung würden beispielsweise die letzten Minuten der aufgezeichneten Daten verwendet werden, um die aktuelle Verkehrssituation zu repräsentieren. Durch die Betrachtung eines geringfügig größeren Zeitfensters, z. B. einer Stunde, und der interaktiven Verschiebung dieses Zeitfensters über einen oder mehrere Tage, ließen sich dagegen beispielsweise Verkehrsströme sichtbar machen, welche sich in Abhängigkeit von der Tageszeit verändern.

Einen anderen Anwendungsfall stellt die retrospektive Analyse dar, in dem historische Daten untersucht werden. Hierbei können durch die Auswahl des Zeitraums temporale Fragestellungen untersucht werden, z. B. Abhängigkeiten bestimmter Phänomene vom Zeitpunkt (wie dem Vergleich zwischen der Situation am Tag und in der Nacht) oder die Veränderung der Gesamtsituation über die Zeit.

6. Anwendungsfälle und Einsatzgebiete

Die hier vorgestellten Verfahren wurden insbesondere im Hinblick auf den Entwurf entscheidungsunterstützender Systeme im Bereich der Verkehrsüberwachung entwickelt. Dichte- und Distanzanalysen stellen dabei zweidimensionale Karten zur

Verfügung, die beispielsweise in einem Geoinformationssystem als Kartenlayer verwendet werden können. Während die erzeugten Dichtekarten zweidimensional sind, können die zugrunde liegenden Bewegungstrajektorien jedoch auch dreidimensionale Informationen enthalten, daher muss untersucht werden, inwiefern sich die Methoden für einzelne Anwendungsgebiete eignen.

Im Falle des Schiffsverkehrs liegen zweidimensionale Bewegungsdaten vor, welche sich direkt für die Berechnung von Dichte- und Distanzkarten eignen. Bereits im Straßenverkehr, welcher im Wesentlichen ebenfalls als zweidimensional angesehen werden kann, gibt es jedoch auch dreidimensionale Informationen, welche unter Umständen falsch interpretiert werden könnten. Beispielsweise stellen Autobahnkreuze oder Über- und Unterführungen von Straßen einen Fall dar, in dem Verkehrswege faktisch voneinander getrennt sind, jedoch auf den gleichen Geokoordinaten verortet werden. Verkehr, der auf zwei sich in dieser Weise kreuzenden Straßen stattdessen, würde daher zusammengerechnet werden und so falsche Ergebnisse erzeugen.

Im Unterschied dazu stellen Bewegungsdaten aus dem Flugverkehr zwar grundsätzlich dreidimensionale Informationen dar, jedoch lassen sich die zu untersuchenden Phänomene im Wesentlichen auch auf zwei Dimensionen reduzieren. Bei der Untersuchung von lokalen Belastungen durch den Flugverkehr, z. B. im Hinblick auf Schallemissionen, spielt die Höhe nur insofern eine Rolle, als dass Flugzeuge ab einer bestimmten Höhe weniger oder gar keinen Einfluss mehr auf die Geräuschentwicklung am Boden haben. Hier bietet sich daher eine Gewichtung des Einflussfaktors in Abhängigkeit von der Höhe an. Zur Erzeugung von Distanzkarten aus Flugbewegungen ist ebenfalls eine Projektion der Flugbewegungen auf 2D-Koordinaten sinnvoll, um eine geografische Zuordnung vornehmen zu können. Bei der Berechnung der räumlichen Distanz muss in diesem Falle jedoch die Höhe der Objekte mit einbezogen werden. Bei der Betrachtung von Flugverkehrswegen kann aufgrund der Organisation des Flugbetriebs eine Unterteilung in zwei unabhängige Karten für den oberen und unteren Luftraum vorgenommen werden. In anderen Fällen kann von vornherein anstelle der geografischen eine topologische Sicht, z. B. auf Flugrouten oder Straßen, sinnvoll sein.

Ein weiteres Einsatzgebiet für automatisch berechnete Dichtekarten kann auch in der Beobachtung von Tierbewegungen, z. B. von Zugvögeln oder Fischeschwärmen, liegen. Insbesondere die Aggregation von Dichtekarten über einen langen Zeitraum kann hier beispielsweise Rückschlüsse auf bevorzugte Aufenthaltsorte, Rückzuggebiete oder wiederholt besuchte Nistgebiete zulassen. Neben der manuellen Exploration der Daten kann auch eine automatisierte Generierung und Auswertung der Dichtekarten sinnvoll sein, welche z. B. automatisierte Jahresauswertungen berechnet und miteinander vergleicht oder mithilfe von Bildanalyseverfahren geografische Regionen ermittelt, welche eine besonders hohe Dichte aufweisen. Diese Analysen können als Ausgangspunkt für weitere (manuelle) Untersuchungen durch die Forscher herangezogen werden.

7. Zusammenfassung und Schlussfolgerungen

Die vorgestellten Visualisierungsverfahren ermöglichen einen direkten Zugang zu massiven raumzeitlichen Bewegungsdaten und ihren Charakteristiken und folgen dabei einem explorativen Ansatz. Es werden Dichtekarten verwendet, die auf der Aggregation von Daten beruhen, um raumzeitliche Phänomene zu extrahieren und zu

repräsentieren. Durch die Auswahl der betrachteten räumlichen und zeitlichen Regionen kann ein Datensatz dabei interaktiv exploriert und analysiert werden.

Durch die Aggregation von Daten über die Zeit werden zeitliche Verläufe nicht explizit dargestellt, wie dies beispielsweise bei der Verwendung eines Space-Time-Cubes [7] [8] der Fall wäre, sondern werden über den gewählten Zeitraum zusammengefasst. Dies ermöglicht es, durch eine interaktive Anpassung des Zeitraums raumzeitliche Phänomene wie Häufigkeit und Dichte von Objekten zu erfassen und so eine andere Form der raumzeitlichen Exploration [9] zu erreichen.

Die Visualisierung und Analyse aggregierter Daten über Zeit und Raum ermöglicht es zum einen, die zeitliche Entwicklung von raumzeitlichen Phänomenen interaktiv zu erkunden und dabei nicht nur zeitliche Momentaufnahmen zu betrachten, sondern statistische Operationen wie Maximal-, Minimal- oder Mittelwerte zu berechnen und deren zeitliche Veränderungen zu betrachten. Zum anderen können so Abhängigkeiten der untersuchten Phänomene, z. B. starke Belastungssituationen oder räumliche Engpässe, von den gewählten Eingabeparametern wie Raum, Zeit oder anderen Parametern der Eingabedaten beobachtet werden.

Die vorgestellten Verfahren werden derzeit innerhalb eines Frameworks zur interaktiven Exploration und Analyse massiver raumzeitlicher Daten implementiert und anhand konkreter Anwendungsfälle aus dem Bereich des Flugverkehrs angewendet.

Danksagung

Diese Arbeit wird im Rahmen des Verbundprojektes PROGRESS (www.earth-in-progress.de) und des BMBF-Förderprogramms „Spitzenforschung und Innovation in den Neuen Ländern“ vom Bundesministerium für Bildung und Forschung der Bundesrepublik Deutschland gefördert. Ein besonderer Dank gilt außerdem der Deutschen Flugsicherung (DFS), insbesondere Herrn Weber, Herrn Dr. Renz und Herrn Wichmann, für die fachliche Zusammenarbeit.

Referenzen

- [1] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, F. Giannotti, *Interactive visual clustering of large collections of trajectories*, Proceedings of IEEE VAST (2009), 3-10
- [2] G. Andrienko, N. Andrienko, *Interactive cluster analysis of diverse types of spatiotemporal data*, SIGKDD Explor. Newsl. **11**, 2 (2010), 19-28
- [3] U. Demšar, K. Verrantaus, *Space-time density of trajectories: exploring spatio-temporal patterns in movement data*, International Journal of Geographical Information Science **24**, 10 (2010), 1527-1542
- [4] N. Willems, H. Van De Wetering, J. J. Van Wijk, *Visualization of vessel movements*, Computer Graphics Forum **28** (2009), 959-966
- [5] N. Andrienko, G. Andrienko, P. Gatalisky, *Exploratory spatio-temporal visualization: an analytical review*, Journal of Visual Languages & Computing **14**, 6 (2003), 503-541
- [6] J. Trame and C. Keßler, *Exploring the Lineage of Volunteered Geographic Information with Heat Maps*, GeoViz (2011)
- [7] M. Kraak, *The space-time cube revisited from a geovisualization perspective*, Proceedings of the 21st International Cartographic Conference (2003)
- [8] P. O. Kristensson, N. Dahlbäck, D. Anundi, M. Björnstad, H. Gillberg, J. Haraldsson, I. Mårtensson, M. Nordvall, J. Ståhl, *An Evaluation of Space Time Cube Representation of Spatiotemporal Patterns*, IEEE Transactions on Visualization and Computer Graphics **15**, 4 (2009), 696-702
- [9] N. Andrienko, G. Andrienko, P. Gatalisky, *Visual Data Exploration using Space-Time Cube*, Proceedings of the 21st International Cartographic Conference (2003)